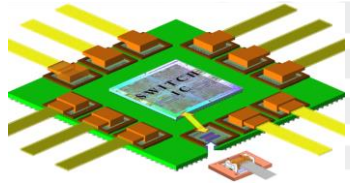
Co-Packaging on Organic Laminates: MOTION Phase 2

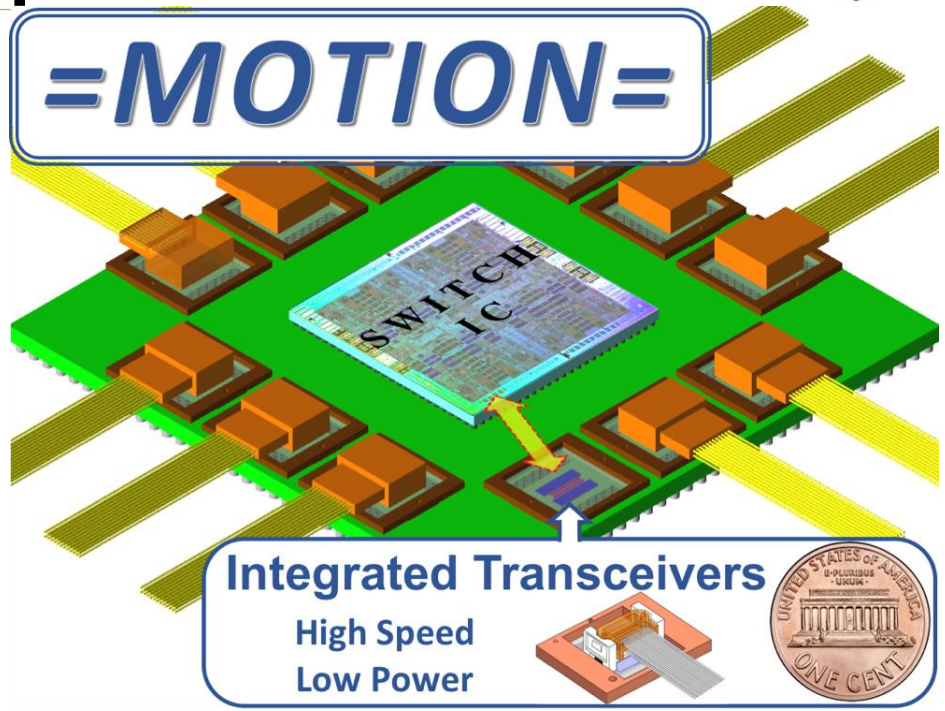
ARPA-E ENLITENED Kickoff Meeting 1/13/2021

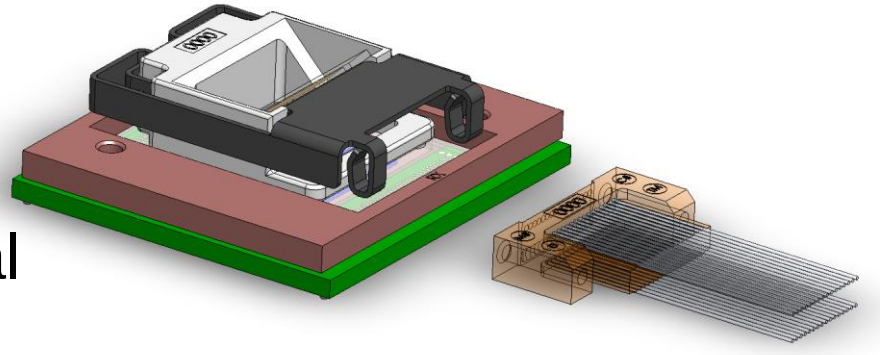
Daniel Kuchta, PI, IBM Research

MOTION Phase 1
Co-packaging for CPU/GPU High-level Specifications



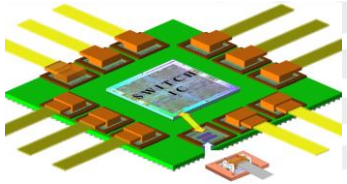
- ARPA-E Sponsored Project on co-packaging
- IBM and Finisar collaboration
- 56GBd NRZ; BER tested to <1E-12 pre-FEC
- 0ºC to 70ºC Case
- 6dB (electrical) link budget (XSR-like)
- 2 dB optical link margin (30m w/connectors)
- Solderable onto ASIC 1st level substrate
- < 4 pJ/bit (3.2W, 16 channels)
- W:13mm x D:13mm x H:4mm
- 25¢/Gb/s



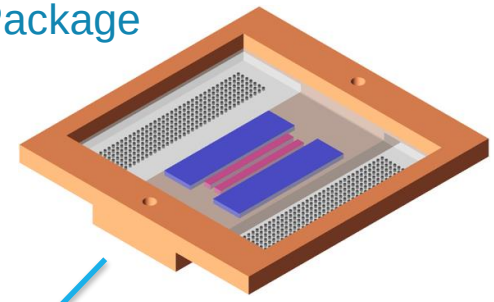


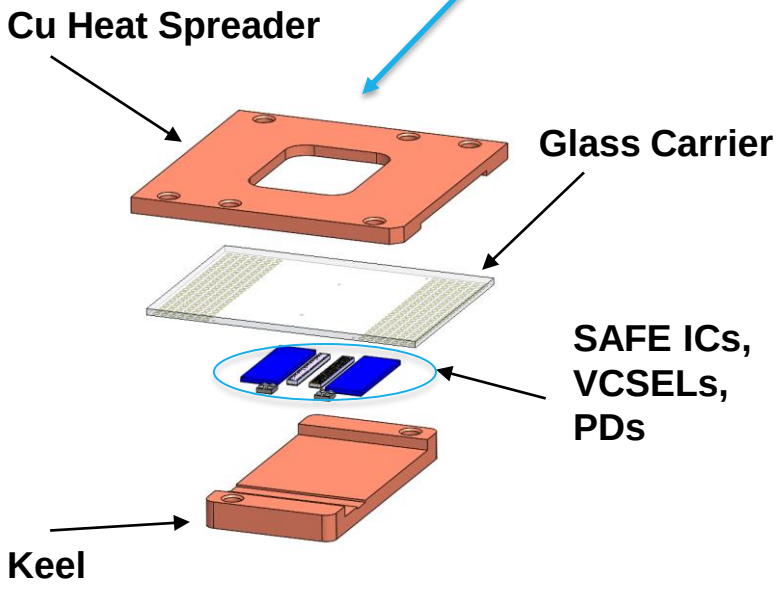
=MOTION=: Multi-wavelength Optical Transceivers Integrated on Node

MOTION Transceiver Package Overview

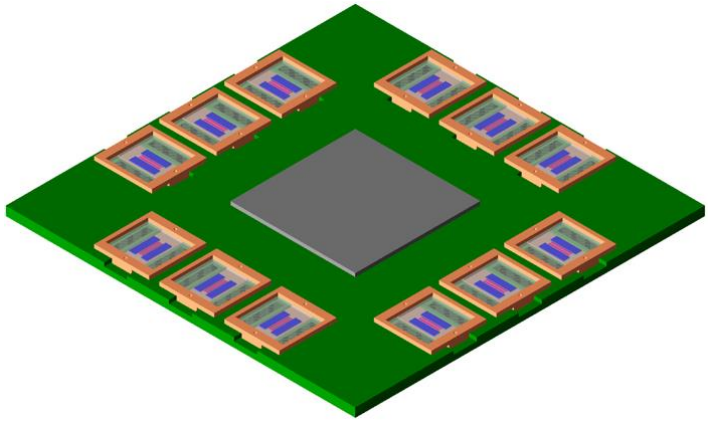


Chip-Scale Optical Package (CSOP)

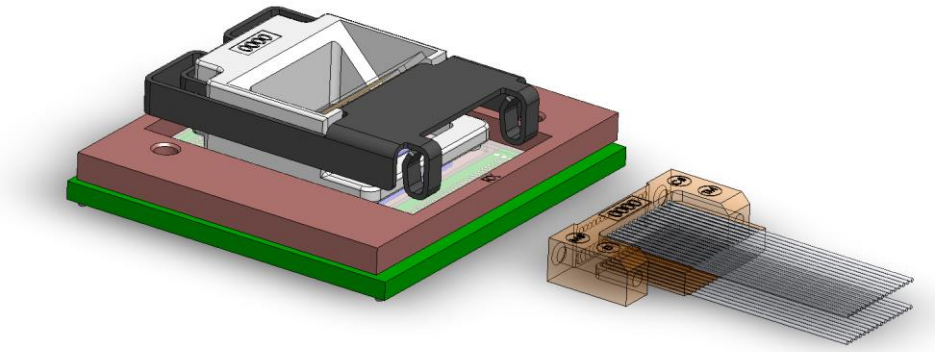




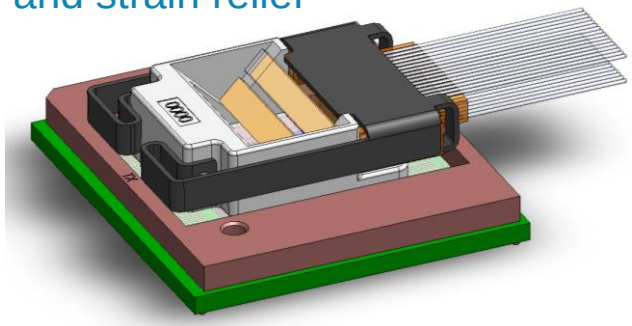
MOTION Vision: Multi-Component Carrier with CSOP for high speed I/O



Final Assembly with lens and clip attached

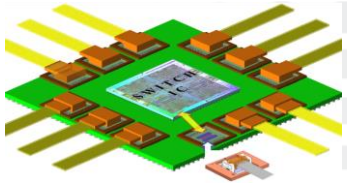


Fully Assembled with fiber cable and strain relief

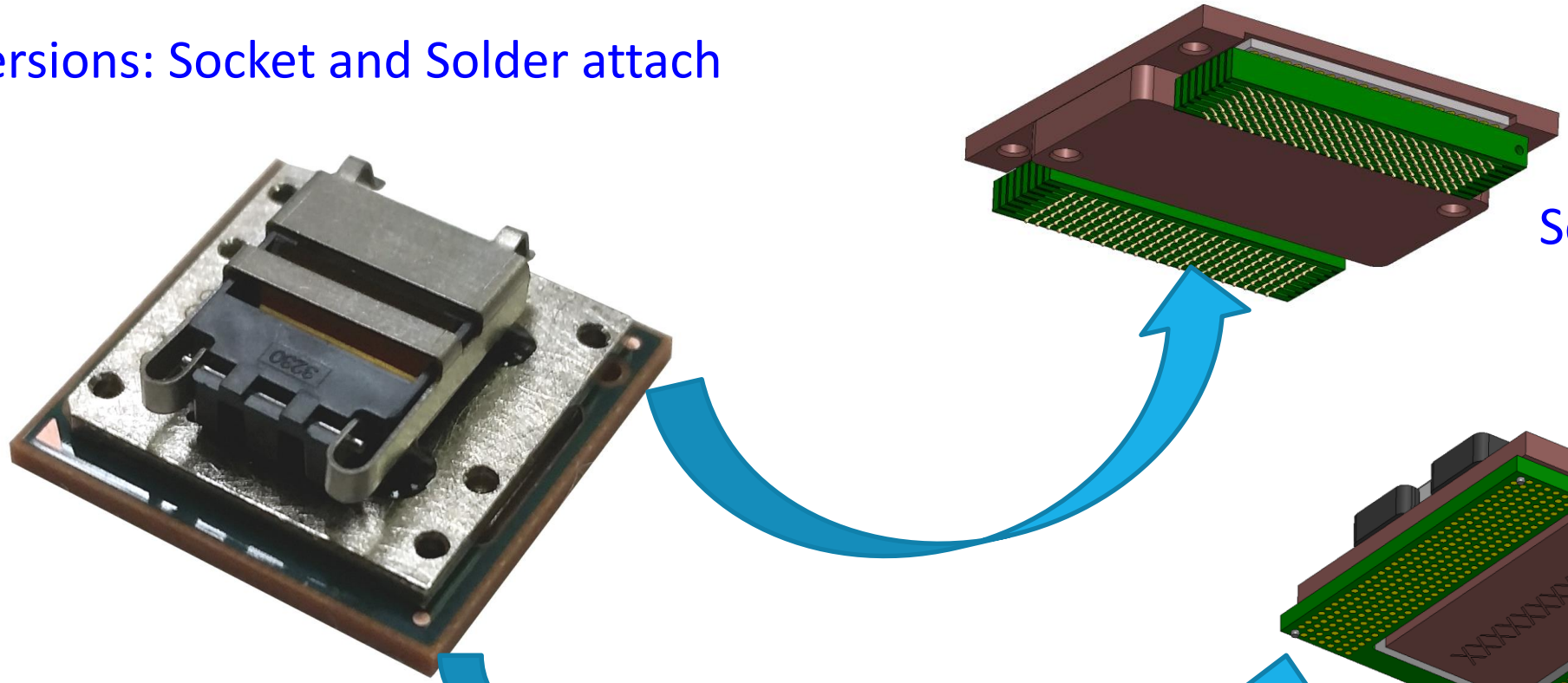


4mm total height

A 13mm x 13mm package with socket insert or interposer



Two Versions: Socket and Solder attach

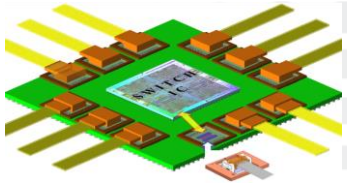


Socket version

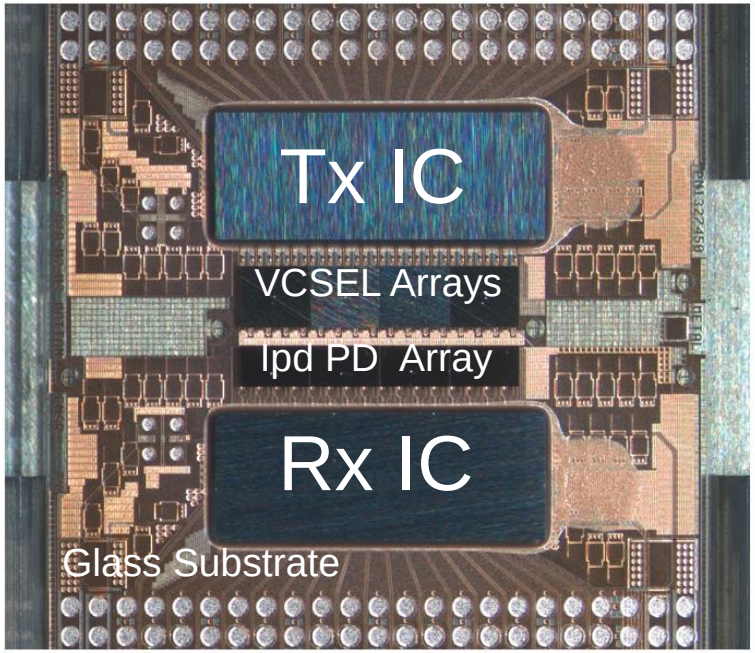
Solder attach version

Sockets may be helpful but incur cost as well as signal & area loss!

Pictures of the completed MOTION modules

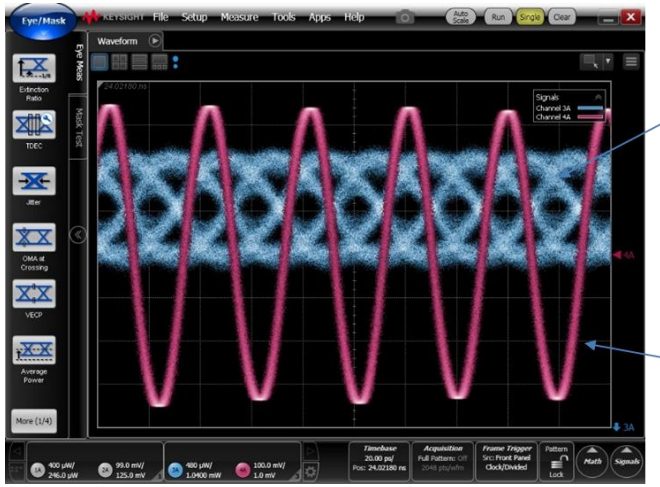


13mm

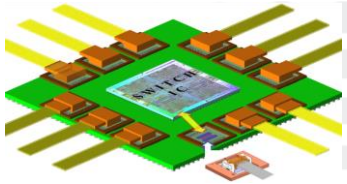


ICs+OEs on glass carrier

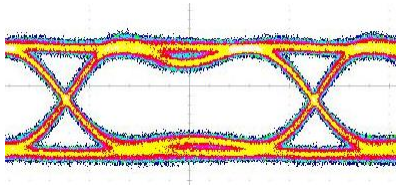
50Gbps NRZ data



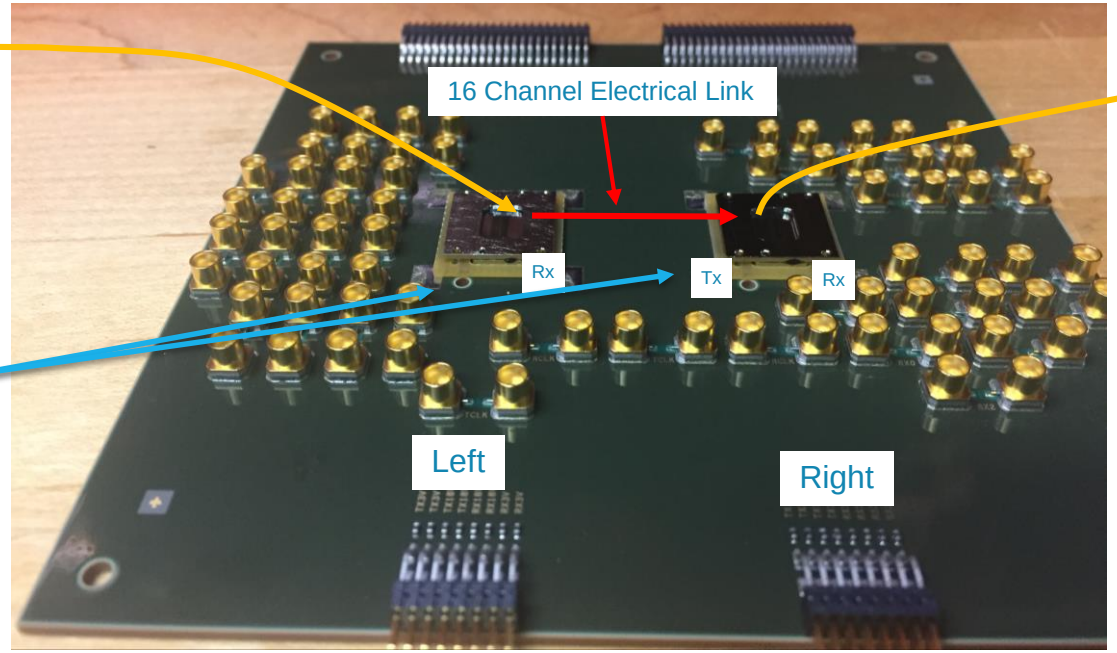
MOTION to MOTION optical & electrical link testing



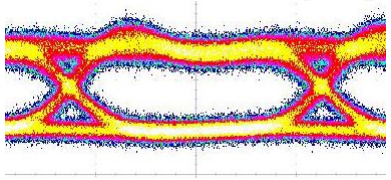
Source Tx

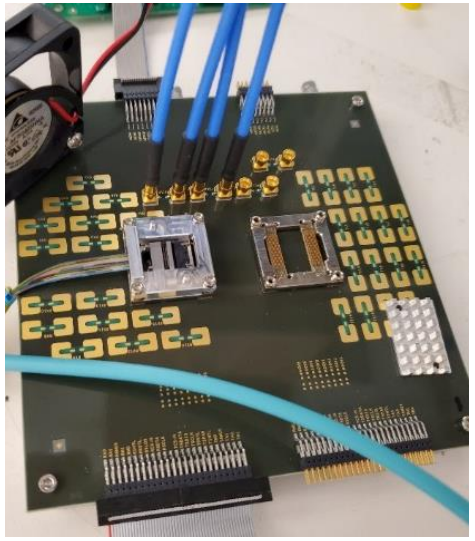


2 MOTION modules soldered to laminate



High Speed Detector (D25)

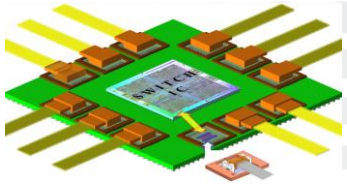




Laminate with MOTION sockets

- MOTION has a low energy electrical interface designed for on-laminate links.
- Unique to MOTION is a combination optical and electrical link test bed for exercising the electrical link ahead of any high speed ASIC that supports it's bandwidth

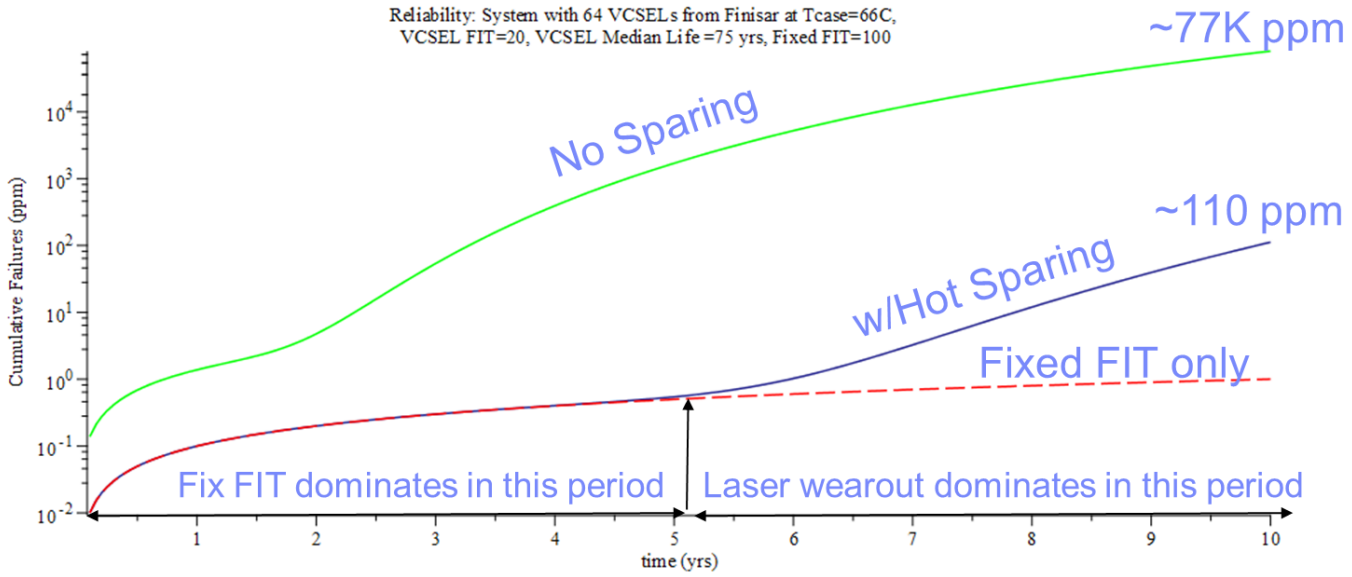
MOTION Approach to Reliability: 2-to-1 Sparing



- Lack of field replacement drives stringent reliability requirements
- Laser wearout dominates: Sparing is desired

- MOTION has 2:1 laser redundancy on every channel
- Simulation shows ~1000x improvement in reliability at the end of 10 years of service →

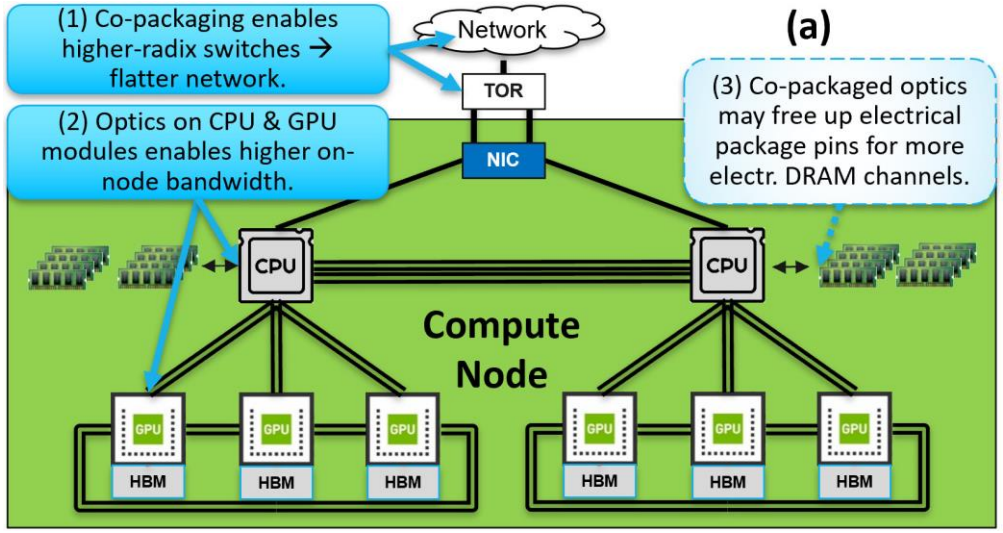
Reliability Simulation for 4 CSOPs on MCC



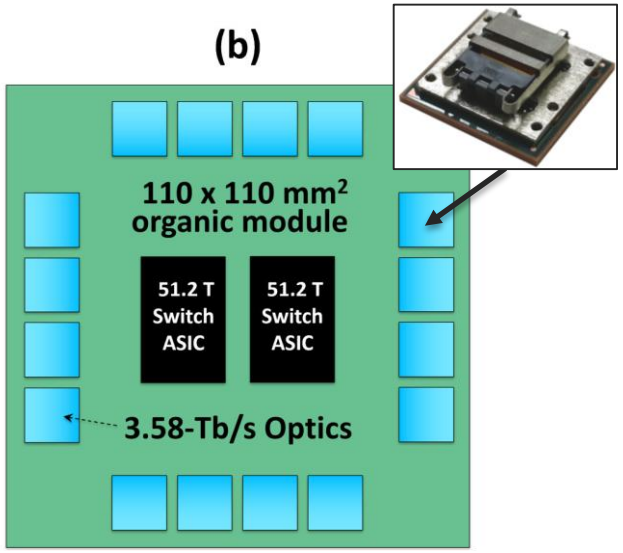
Assumed Parameters:
VCSEL MTTF = 75 yrs, VCSEL FIT = 20
Fixed FIT = 50 FIT per module
Ibias = 9 mA

Fixed FIT: package components that can not be spared
= Sum of everything with a non-zero FIT

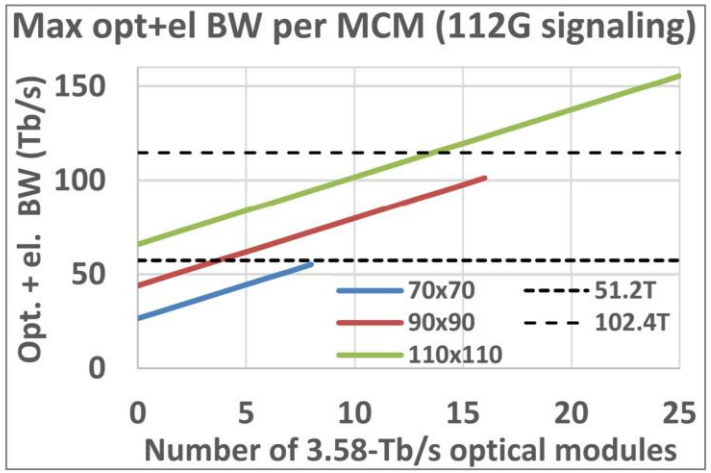
M&S: How much additional bandwidth can MOTION provide?



Possible insertion points of co-packaged optics on switches, CPUs or GPUs



Example of co-packaged optics enabling a 102.4-Tb/s switch with 16 13x13 mm² modules



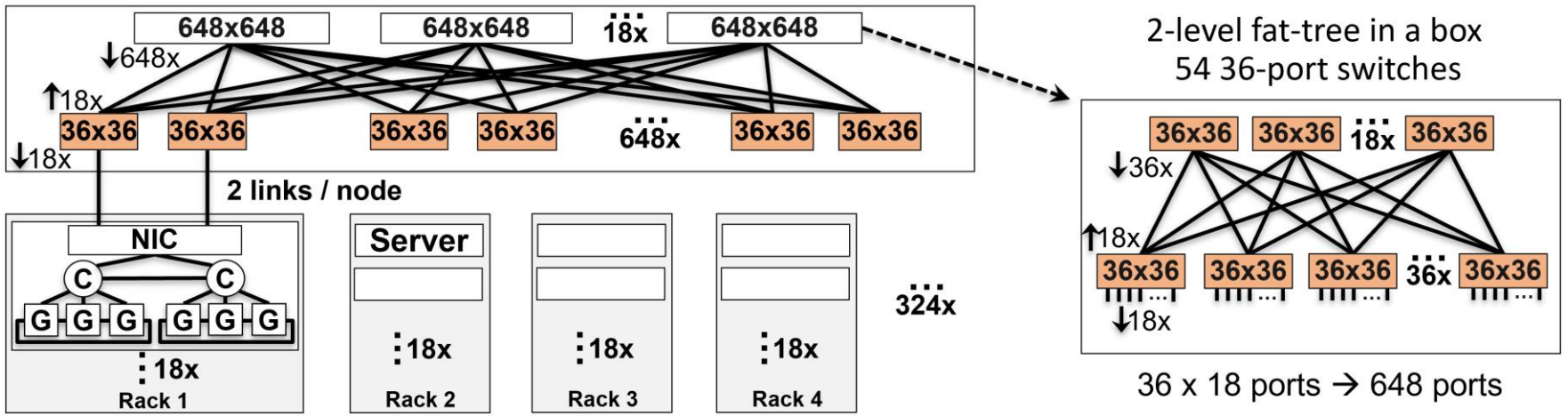
Max switch BW for an up to 40% fill factor for the free carrier area

- **P. Maniotis, et. al., *Scaling HPC Networks with Co-packaged optics*, OFC 2020, San Diego**
  - CPO-enabled 51.2-Tb/s switch module with 128 400-Gbps ports
  - Vs Summit-like tech. (for 3 switch layers): (a) 2.8x more end points, (b) 11.2x higher bisection BW, (c) 21% fewer switches

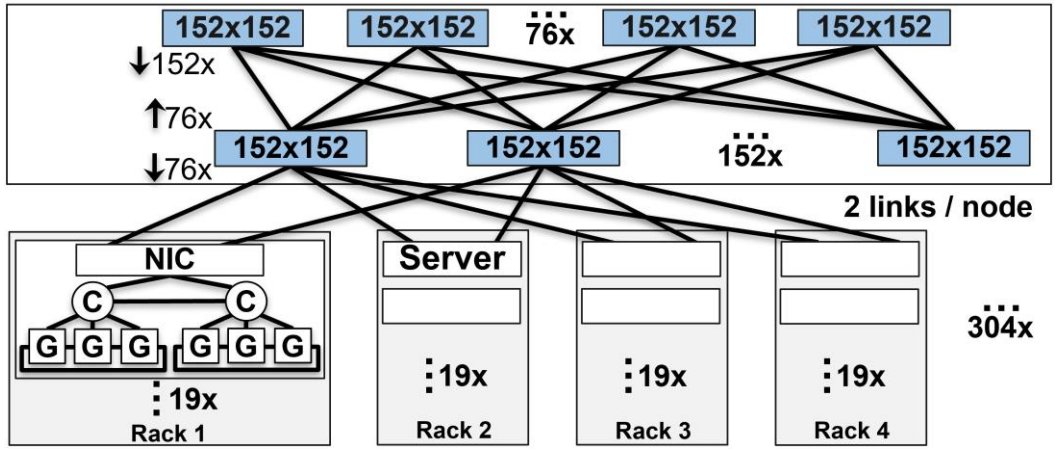
- **P. Maniotis, et. al., *Toward lower-diameter large-scale HPC and data center networks with co-packaged optics*, JOCN, Jan. 2021**
  - CPO-enabled 60.8-Tb/s switch module with 152 400-Gbps ports
  - Vs Summit-like tech. (for similarly sized net.): (a) 3 vs 5 max hops (2 vs 3 switch layers), (b) 4x higher bisection BW, (c) 86% fewer switches

M&S: The benefits of higher-radix switches enabled by MOTION

Baseline network – 11,664 end points – 1,620 36-port switches



MOTION network – 11,552 end points – 228 152-port switches

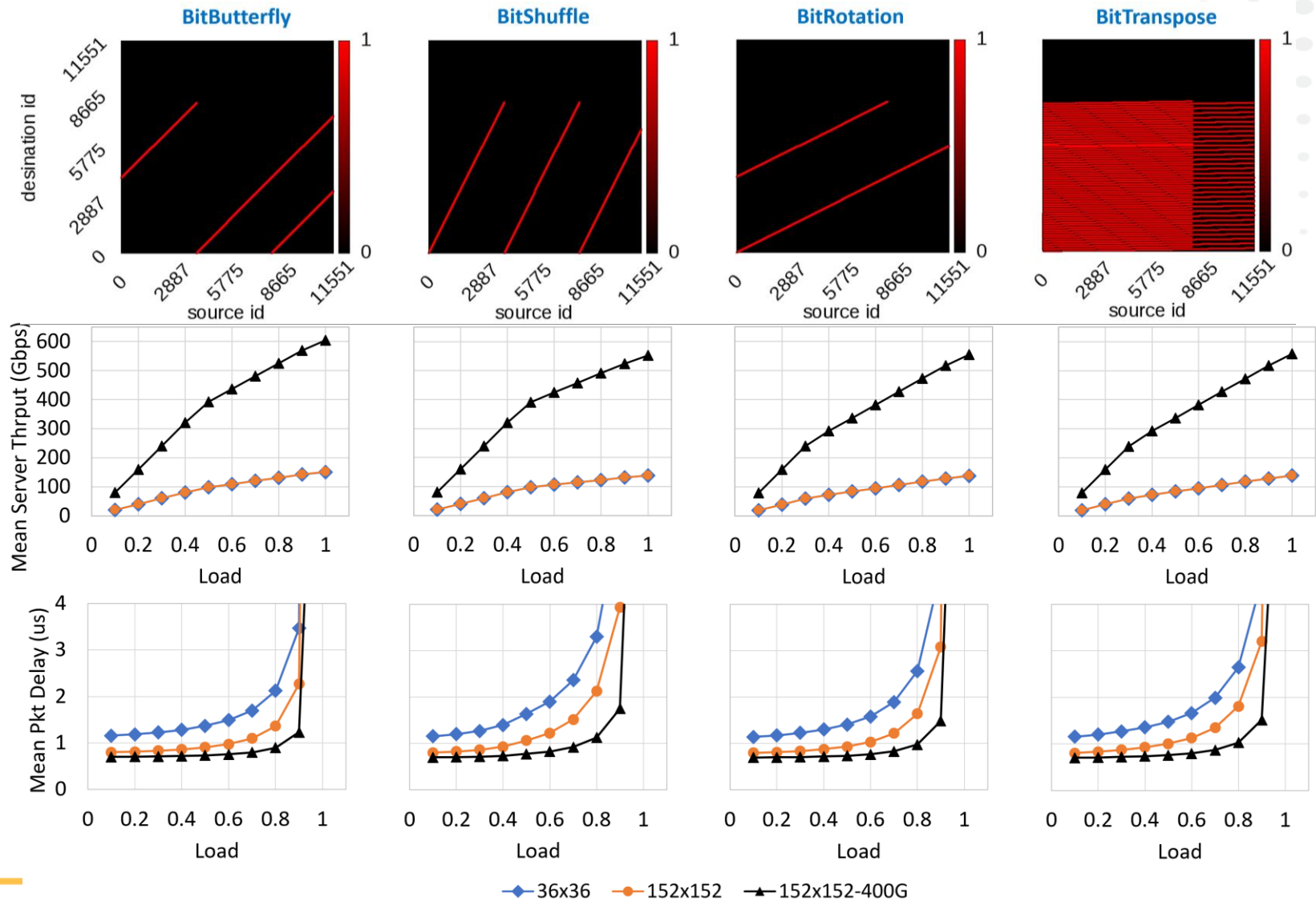


- 4x bisection bandwidth
- Flatter network → (a) 3 vs 5 max hops; (b) Reduced latency; (c) Less network contention
- 86% fewer switches → (a) Reduced cost; (b) Reduced energy consumption; (c) Easier management & administration

M&S Performance analysis: 4 synthetic benchmarks w/ hotspots
(MOTION-2 → to extend w/ HPC/Datacenter benchmarks of interest)

Venus discrete-event network simulator

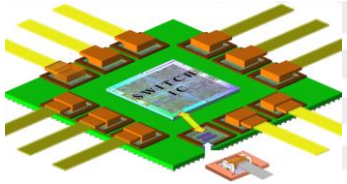
Simulation results	
Generator	
Packet size	1500 B
Generation distribution	Bernoulli
Data rate	100/400 Gbps
Load	[0.1-1]
Adapter/Switch	
Technology	InfiniBand
Data rate per link	100/400 Gbps
Delay	100 ns
Switch buffer / port	128 KB



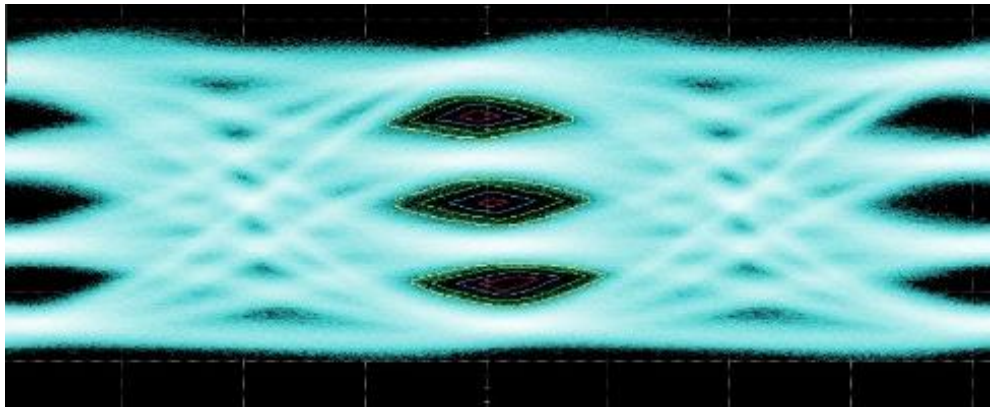
- 152x152: Same bisection BW → same throughput
  30-36% mean pkt delay improvement
- 152x152-400G: 4x higher throughput/server
  40-70% mean pkt delay improvement

arpa·e CHANGING WHAT'S POSSIBLE

Co-Packaging on Organic Laminates: MOTION Phase 2

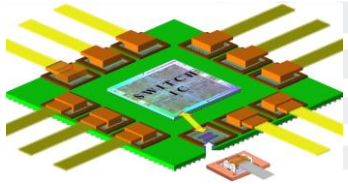


- **ARPA-E (U.S. Department of Energy) sponsored project, Phase 2: 2 years**
  - IBM and Finisar Inc. (now II-VI Inc.)
- **Target specifications**
  - **Optimized Electrical Interface for organic on-laminate channels**
  - **Optical Interface: 32 channels @ 112G PAM-4, 16 fibers, 2 wavelengths**
  - **< 2 pJ/bit (7W, 32 channels)**
  - 0ºC to 70ºC Case
  - 6dB (electrical) link budget (XSR-like)
  - 2 dB optical link margin (30 to 50m w/connectors)
    W:13mm x D:13mm x H:4mm
  - Package can withstand reflow onto ASIC 1st level substrate



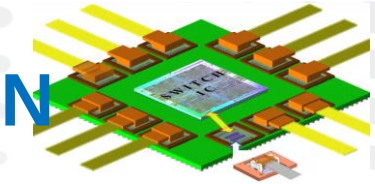
940nm VCSEL @ 112G PAM-4

MOTION Phase 2 Proposed Hardware Changes

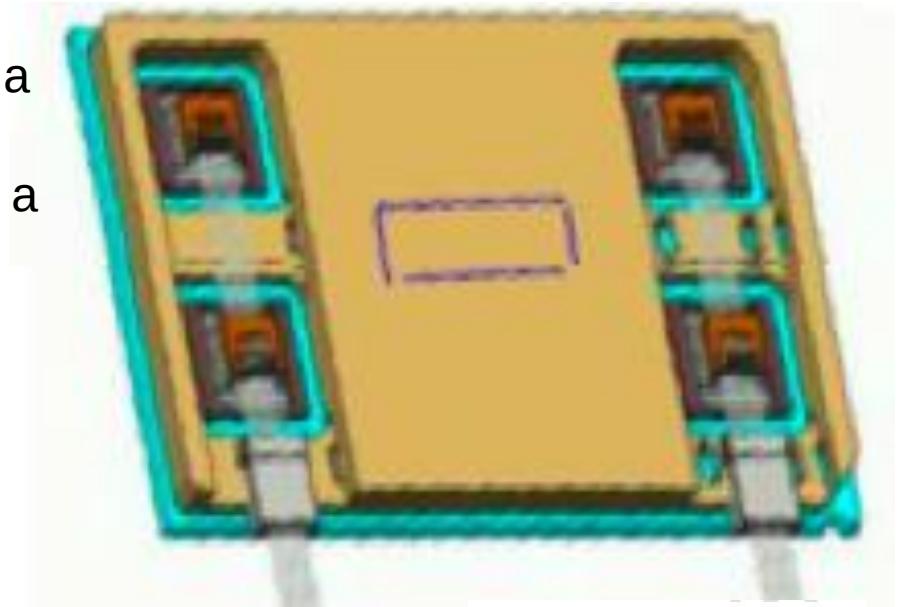


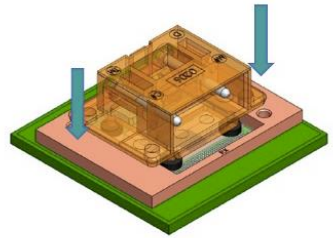
Parameter	Phase 1	Phase 2
Electrical Interface	16 channels @ 56G NRZ	> 32 channels @ tbd_G NRZ/PAM4
IC Technology	SiGe	CMOS
Optical Interface	16 channels @ 56 G NRZ	32 channels @ 112G PAM4
# of Wavelengths	1	2
# of Fibers	16 Tx + 16 Rx	16 Tx + 16 Rx
Fiber Type	50/125 MMF	50/125 MMF
Package I/O Pitch	400um	300um
Glass Carrier Size	13x13mm	13x13mm
Energy consumption	4 pJ/bit	2 pJ/bit
Projected cost	25¢/Gig	TBD but <25¢/Gig
Laminate interface	Soldered or LGA	Soldered or LGA

Phase 2: IBM SYSTEMS Group TECHNOLOGY EVALUATION



- Goal: <u>Assess the technology readiness of co-packaged optics</u>
  - Optical transceivers will be soldered directly on the top surface of a production laminate package
  - Four (4) optical transceivers and one (1) test site die on the top of a single FC-PLGA laminate, assembled with a thermal lid
  - Two evaluation cycles with positive results would result in a recommendation that this technology could move into a productization phase
  - Socketed optical transceivers will be evaluated through modeling
  - Focus on the thermal & mechanical robustness of packaging
- Evaluation challenges:
  - Assembly Processing
  - Package Reliability
  - Thermal Performance
  - Modelling





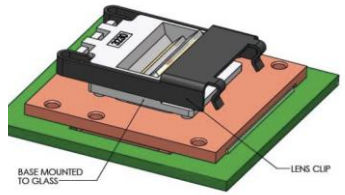
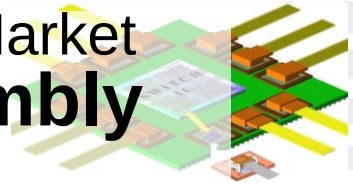


Figure 2 – Configuration with Soldered Interposer

Demonstrating a viable path to system integration is necessary before this technology can be included in a product plan, qualification, and commercial offering.

MOTION2, 3.2T (32x 112G-PAM4) Co-packaged Optical Assembly

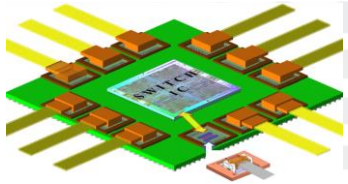


- The MOTION2, 3.2T Co-packaged Optical Assembly(COA) is targeted at high-bandwidth, optical-interconnect applications which value:
  - High bandwidth density per square-millimeter
  - Protocol agnostic capability up to 112Gbps-PAM4/channel on 32-duplex channels
  - Low power-consumption
  - Low latency applications by utilizing NRZ encoding up to 56Gbps without FEC

- Four market segments Identified and Pursuing COA applications:
  - Datacenter Networking
  - High Performance Computing & AI-Deep Learning interconnect (Massively Parallel Processing)
  - Metro-access Edge Compute Equipment for edge-datacenters with hyper-converged or disaggregated resource pools
  - High-performance FPGA interconnects serving Aerospace, High-resolution Imaging, & future accelerator technologies

- Key challenges to commercialization are customer timeline for releasing applications, obtaining funding to support advanced development efforts, and defining a new COA/ASIC assembly, & final-test, supply-chain model to replace the pluggable transceivers model of the past two decades

Challenges for Co-packaged Optics



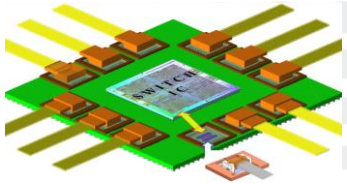
- Reliability
- Field Replacement/Serviceability
  - Fail-in-place Strategy?
- Yield
  - Who is responsibility for final yield?
- Assembly
  - Who does what and when?
- Standards and/or MSAs
  - Minimum Time for Standards seems to be > 2 years
  - Proprietary solutions likely to emerge first
- Compatible Technologies
  - MMF or SMF
- Field Upgradeable Firmware!

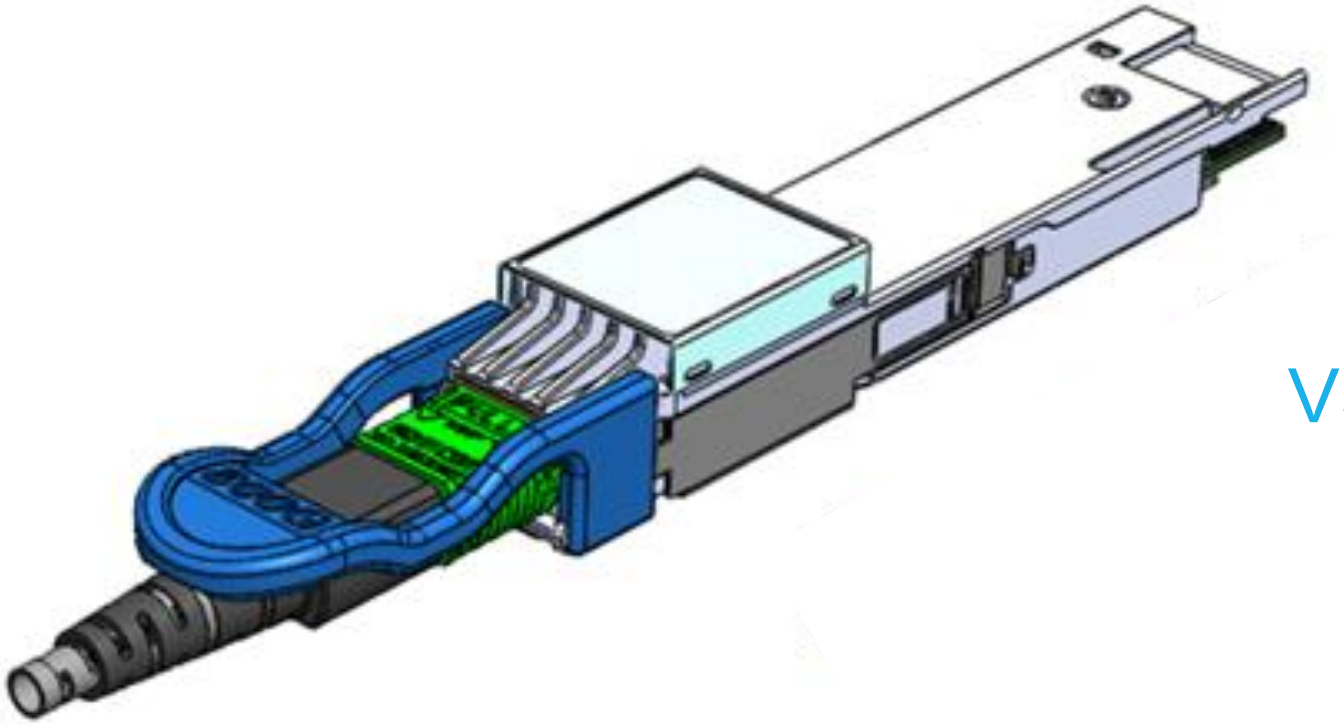
Ruminations about SiPh Co-Packaging

- Multiple unproven technologies introduced simultaneously to the heart of a system is risky.
- External high power lasers with PM fiber and internal Y cables with high fiber count: Unlikely to be cost competitive

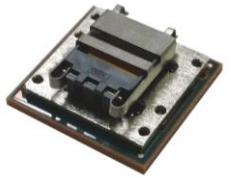
→SiPh still can't beat the cost of VCSELs+MMF

Size Comparison of QSFP-DD (DR8) and MOTION1





VS



To scale; Both provide 800Gbps Bi-Directional Bandwidth
but only one of these will reach 25¢/Gig…..